Carnegie Foundation for the Advancement of Teaching

STATWAY™ INSTRUCTOR NOTES

Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line

INSTRUCTOR SPECIFIC MATERIAL IS INDENTED AND APPEARS IN GREY

ESTIMATED TIME

50 minutes

MATERIALS REQUIRED

None, although you might wish to have an overhead or electronic display of the graphs in the lesson.

BRIEF DESCRIPTION

This lesson lays the foundations for understanding the meaning of the slope and the y-intercept using the LSR line.

LEARNING GOALS

Students will understand that:

- A line is determined by two parameters: its slope and y-intercept.
- The slope is described by a ratio of the change in y relative to the change in x and can be interpreted as the predicted change in y associated with a one-unit change in x .
- The y-intercept is the y-value when $x = 0$. Frequently, this point does not have meaning in the context of the data because $x = 0$ lies outside the range of the data. Nevertheless, the initial value is a part of the algebraic structure of the equation of the line.
- The slope of the least squares regression line may be affected by outliers, particularly outliers with extreme values for the predictor variable.

Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line

Students will be able to:

- Interpret (in context) the slope of the least squares line.
- Interpret (in context) the y-intercept of the least squares line, if appropriate.
- Identify observations that are outliers and observations that are potentially influential given a scatterplot.

INTRODUCTION

Instead of the usual approach in algebra of presenting the slope formula and then equations of lines in the form $y = mx + b$, you will discuss the equation of the line from the perspective of predicting y. The concepts of slope and y-intercept are developed as a way to predict y for a given x. In Module 12, you will revisit linear functions from the more traditional algebraic perspective.

No extensive introduction is required for this lesson. Tell students that the lesson's general goal is to figure out what the numbers in the equation of a line tell them. The first task in the lesson reviews concepts related to the least squares line. Later, tasks focus on understanding and interpreting the numbers in the equation of the line.

S STUDENT MATERIAL IS NOT INDENTED AND APPEARS IN BLACK

TRY THESE

Investigating the Line

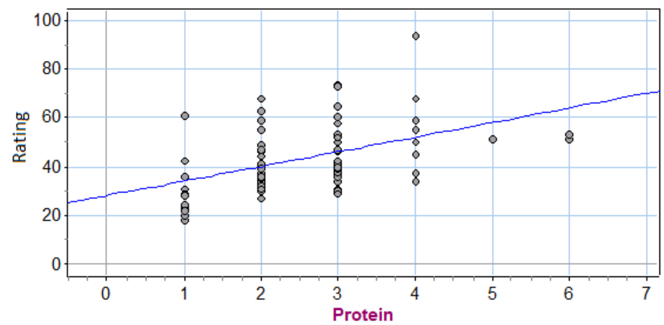
1 Here you return to the cereal data. Begin with a few questions to review what you know about the least squares line. Here is the least squares line that predicts *Consumer Reports* ratings based on the amount of protein (in grams) in a serving:

- Predicted rating = 28 + 6(protein).
- The correlation coefficient is 0.48.

Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line

A Use the least squares regression line to predict the rating for a cereal containing 2 grams of protein in a serving.

> **Answer**: 40.



B There are two cereals with 6 grams of protein in a serving. Is the predicted rating from the least squares regression line too high or too low for these cereals?

> **Answer:** In the middle.

C What does the phrase *least squares* tell you about this line?

> **Answers will vary.** This is the line that minimizes the squared errors. By this criterion, it is the line that best fits the data. It is the best summary of the relationship between protein and ratings.

D Which of these statements would you use to describe protein as a predictor?

- Protein is a very accurate predictor of *Consumer Reports* ratings (errors within a few rating points would be typical).
- Protein is not a very accurate predictor of ratings (errors as large as 10 rating points would not be surprising)

> **Answer:** The second option is more reasonable due to the large variability in ratings for cereals with the same amounts of protein. Correlation is not strong, 0.48.

x = protein (g/serving)	$\hat{y}$= predicted rating
0	28
1	34
2	40
3	46
4	52
5	58

Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line

2 Now you will focus on understanding what information you get from the numbers in the equation of the line. Here we have the equation of the least squares line and a table of protein amounts and predicted ratings from the least squares line.

predicted rating = 28 + 6(protein)

$$\hat{y} = 28 + 6x$$

How are the 28 and 6 related to the table of values? Be as specific as you can.

> **Note:** Some students may recall the idea of slope and y-intercept, but many will not. If students are using these terms and concepts, reinforce the connections that they are making as you work with them. However, many students will not give precise answers to this question. If students are giving cursory answers, such as “28 is in the table,” prompt them to be more precise with questions such as the following:
>
> - Is the 28 a protein amount or a predicted rating?
> - How much protein is in the cereal if the predicted value is 28?
> - When the predicted rating increases by 6, how much more protein is in the cereal?
> - If you remade the table so that the protein amount increased by 2s, would you still see the +6 pattern? What pattern would you see in the predicted ratings? By 3s?)

3 Here you have the least squares equation for predicting *Consumer Reports* ratings based on the amount of sugar (in grams) in a serving. The equation was used to generate the table of predicted values for some sugar amounts.

predicted rating = 60 – 2.4(sugars)

$$\hat{y} = 60 - 2.4x$$

***x* = sugar (g/serving)**	**$\hat{y}$= predicted rating**
0	60
1	57.6
2	55.2
3	52.8
4	50.4
5	48

How are the 60 and –2.4 related to the table of values? Be as specific as you can.

> **Note:** See previous note on intervening as students work.

Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line

WRAP-UP

Discussion of Question 2

x = protein (g/serving)	$\hat{y}$= predicted rating
0	28
1	34
2	40
3	46
4	52
5	58

x = protein (g/serving)	$\hat{y}$= predicted rating
0	28
2	40
4	52
6	64
8	76
5	58

x = protein (g/serving)	$\hat{y}$= predicted rating
0	28
3	46
6	64
9	82
12	100
5	58

What does the 28 tell you? What does the 6 tell you?

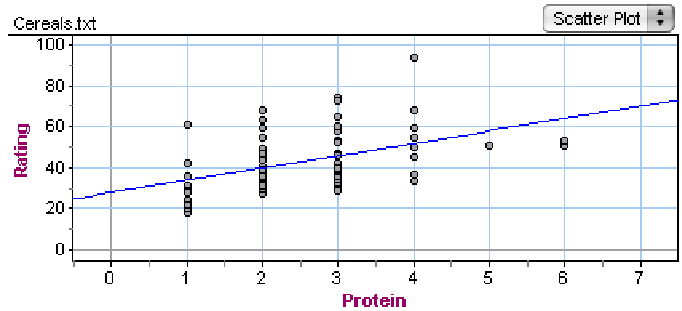
Discuss the Values in the Table: If you think of the line as predicting ratings, you start with an initial rating for a cereal with no protein, which is 28. Then you predict an increase in the rating of 6 for every extra gram of protein put into a serving. So the 6 is the increase in rating *per* gram of protein.

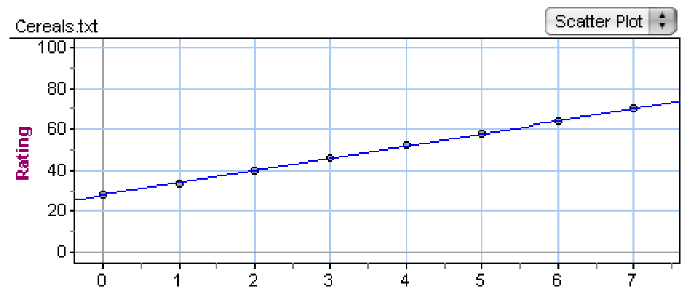
If you use the equation to create a table where you have changed the scale on *x*, the predicted increase in ratings per gram is still the same even though the pattern looks different: increase the rating by 12 for every 2 extra grams of protein, but this is equivalent to saying an increase of 6 in ratings for every gram of protein added to a serving, etc.)

Let's think about what the role that 28 and 6 play when we look at the graph of the line.

Note: First make sure that students understand that the two lines shown are the same line. Describe what dots represent in the two graphs. Dots in the first graph represent data from the cereals. Dots in the second graph are values from the table. These dots are not cereals; they do not represent data. However, they show the pattern predicted by the least squares line.

Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line





Next show that 28 is the *y*-intercept, the *y*-value when *x* = 0. Note that these graphs are drawn with *x* = 0 slightly to the right of the vertical boundary of the graphing window. This is intentional to get students to focus on the fact that *x* = 0 and not the value that happens to mark the left boundary of the graphing window.

In the second graph illustrate the slope triangle for two different pairs of points in the table [with different Δx] to reiterate the point just made in the discussion of the table about a constant predicted increase of 6 in rating score for every additional gram of protein. Then show how any predicted *y*-value can be obtained by starting at the *y*-intercept of 28 and increasing the rating by 6 for each additional gram of protein added to a serving. For example, if *x* = 3 the slope triangle connecting [0, 28] with [3, 46] shows a change in *y* of 18. The predicted *y*-value is 28 plus the change in *y* of 18.

Discussion of Question 3

The least squares line is $\hat{y} = 60 - 2.4x$. What do the numbers 60 and –2.4 tell you?

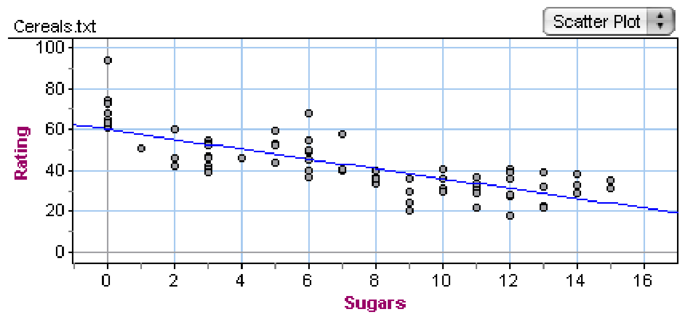
x = Sugar (g/serving)	$\hat{y}$= predicted rating
0	60
1	57.6
2	55.2
3	52.8
4	50.4

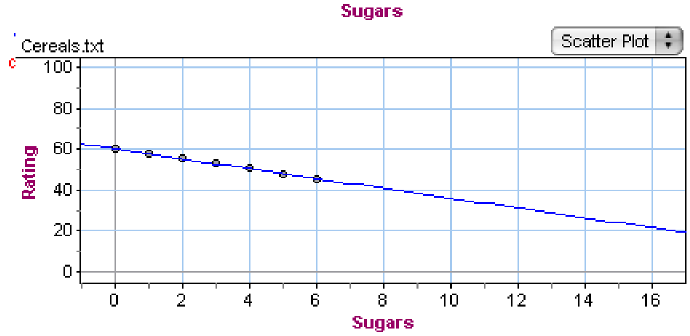
x = Sugar (g/serving)	$\hat{y}$= predicted rating
0	60
2	55.2
4	50.4
6	45.6
8	40.8

x = Sugar (g/serving)	$\hat{y}$= predicted rating
0	60
5	48
10	36
15	24
20	12

Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line

Repeat this same discussion using the sugar-ratings tables and graphs.





If you think of the line as predicting ratings, start with an initial rating for a cereal with no sugar (60), and then decrease the predicted rating by 2.4 for every extra gram of sugar put into a serving. So, the –2.4 is the decrease in rating per gram of sugar.

If you use the equation to create a table where you have changed the scale on *x*, the decrease in ratings per gram is still the same: decrease the predicted rating by 4.8 for every 2 extra grams of sugar, which is equivalent to an increase of 2.4 in predicted ratings for every gram of sugar added to a serving.

It is also worth noting that interpreting the slope in a graph is easier if you think carefully about how to draw the slope triangle. Choose points on the line that fall on or near the intersection of grid lines to make reading the changes in *x* and *y* easier.

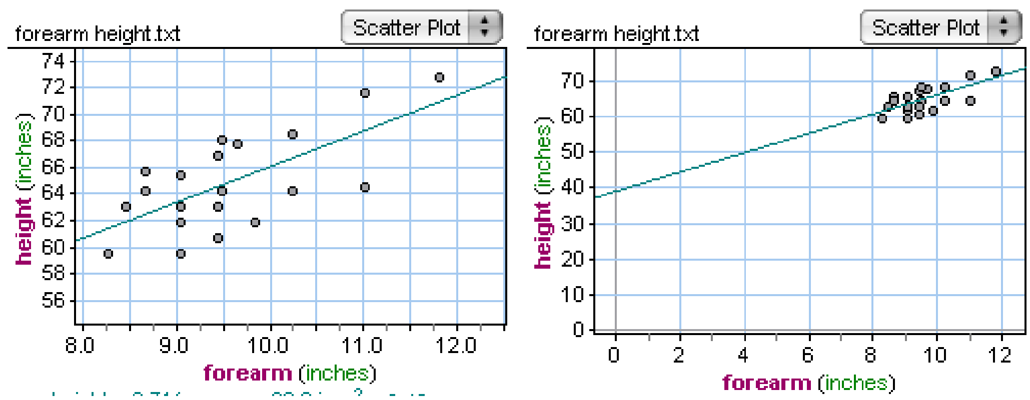
INTRODUCTION

The next tasks require that students identify equations by looking at graphs. Some students may work from an understanding that slope is a measure of steepness. This may lead to some confusion since different scales are used on the axes. So, lines that have a smaller slope relative to other lines may visually appear steeper due to the scaling. Encourage all students to apply the idea of a slope triangle. If students are struggling with determining the slope, help them make smart decisions about how to construct slope triangles by looking for places where the lines cross grid marks. This makes determining the slope easier. Treat these tasks as guided practice and do not hesitate to guide. If necessary, call their attention to the scales on the axes.

Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line

Discussion of a New Least Squares Line for a Different Scenario

Let's return to another data set you used recently, the forearm and height measurements for 21 female college students taking Introductory Statistics at Los Medanos College in Pittsburg, California, in 2009. The equation of the least squares line is $\hat{y} = 2.7x + 39$, where x = forearm length (inches) and $\hat{y}$ =predicted height (inches). What do the numbers 2.7 and 39 tell you?



Note: Here the line is written with the x-term first, which may cause some initial confusion. This is intentional to get students to realize the terms in the equation can be commuted. Typically, statisticians write lines in the form $y = a + bx$.)

Discuss What You Can Tell from the Symbolic Form of the Equation: In this equation, you are predicting height based on forearm length. The equation predicts a height of 39 inches for a forearm length of 0. Obviously, it does not make sense that you can predict a person's height will be 39 inches when that person has a forearm length of 0. You can see from the graph when x = 0 you have extrapolated beyond the range of the data. However, the initial value is an important part of the least squares equation for predicting height, even when the initial value is nonsensical in the context of the data. So, what does the least squares equation tell us? To predict a person's height, start with a height of 39 inches and add 2.7 inches in height for each inch of forearm length.

In the next set of tasks, you will continue to focus on understanding the numbers in the equation of a line.

Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line

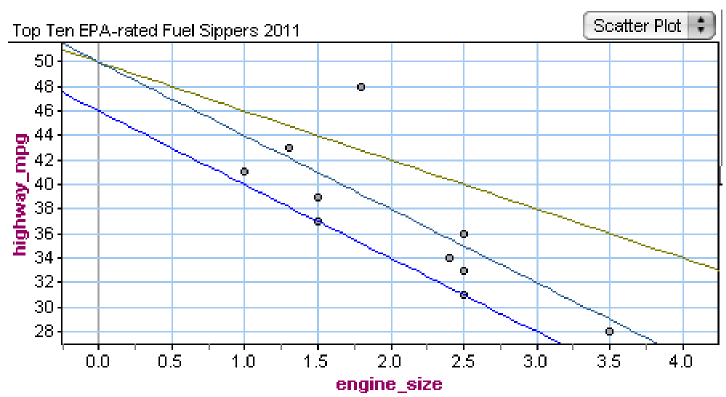
NEXT STEPS

Identifying the Best Fit Line

4 The Environmental Protection Agency picks the 10 most fuel-efficient cars each year. Below is a scatterplot of the highway miles per gallon and the engine size (measured in liters) for the EPA's top 10 for 2011. (Retrieved from www.fueleconomy.gov)

Following are the equations of the three lines shown:

$$\hat{y} = 46 - 6x$$
$$\hat{y} = 50 - 6x$$
$$\hat{y} = 50 - 4x$$



A Identify the *equation* of the line that best fits the data. Briefly explain how you made your decision.

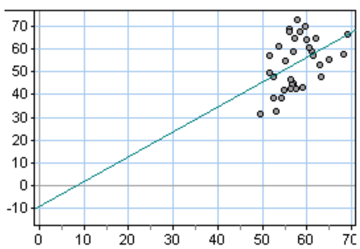
> **Answer:** The middle line does the best job of approximating the data. This line has an initial value of 50 when $x = 0$. A slope triangle through gridlines connects [0, 50] and [1, 44]; when x increases by 1, y decreases by 6. So, the equation is $\hat{y} = 50 - 6x$.

B For the line you chose, describe what the numbers in the equation tells you about engine size, highway miles per gallon, and the relationship between the two for these 10 cars.

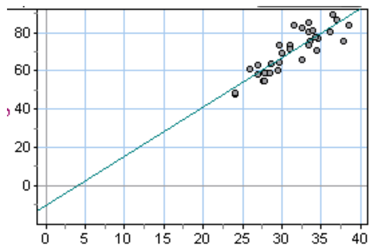
> **Answer:** The predicted highway mileage decreases by 6 for an increase of 1 liter in engine size. The 50 is the y-intercept; it is the highway mileage given by the line for an engine of size 0 liters. This is an example of an initial value that is nonsensical in the context of the data.

Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line

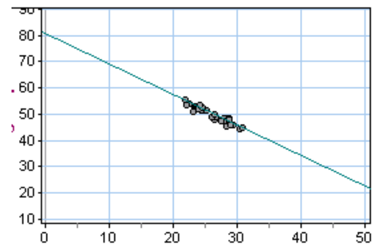
5 Match the graphs to the least squares equations and r-values.



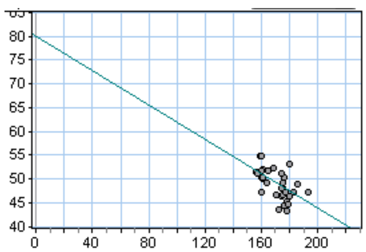
Graph A



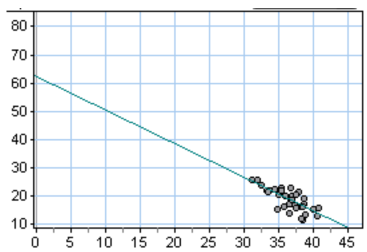
Graph B



Graph C



Graph D



Graph E

Here are r-values to choose from:

	–0.54	–0.73	0.45	–0.95	0.88
Answers:	D	E	A	C	B

Here are equations to choose from:

	$\hat{y} = -10.5 + 2x$	$\hat{y} = 62 + (-1.2)x$	$\hat{y} = -10.5 + 1.1x$	$\hat{y} = 80 + (-1.2)x$	$\hat{y} = 80 + (-0.2)x$
Answers:	B	E	A	C	D

WRAP-UP

Since these two tasks require students to apply concepts previously discussed, just go over the answers and field questions that arise. Return as necessary to the idea of *y*-intercept (when *x* = 0) and slope as the amount *y*-hat changes for each one-unit change in *x*.

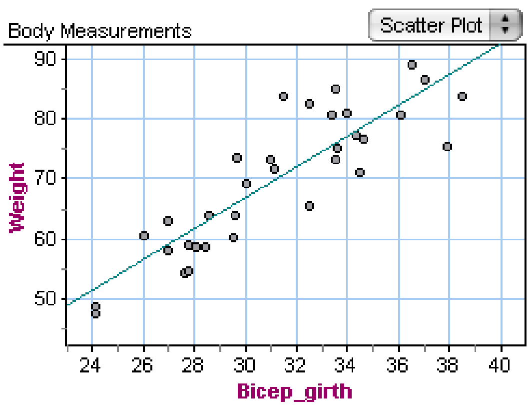
Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line

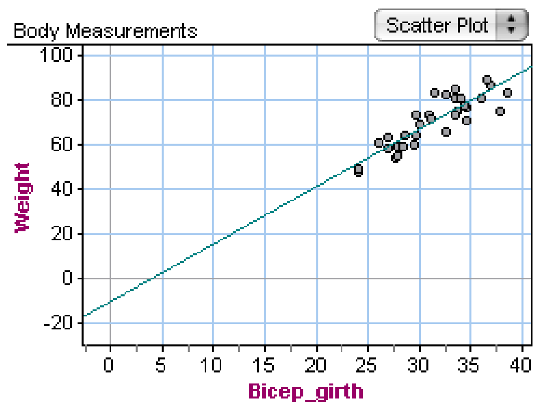
TAKE IT HOME

1 Based on data from 34 adults who exercise regularly, the least squares line for the relationship between bicep girth and weight is

predicted weight = 2.6(bicep girth) – 10.5

where predicted weight is measured in kilograms and bicep girth is measured in centimeters.





A Construct a table or use one of the graphs to explain the meaning of 2.6 in this situation.

> **Answer:** The least squares line predicts an increase of 2.6 kilograms of body weight for each additional centimeter in bicep girth. Even if the wording students use is awkward, their answers should demonstrate, in the context of the variables given, an understanding that the slope is a change in predicted y relative to a change in x.

B In the prediction equation, –10.5 is the initial value when $x = 0$. Does this number have meaning in this scenario? Why or why not?

> **Answer:** The initial value here is nonsensical since it does not make sense that the distance around someone's bicep is 0 centimeters. Obviously, you cannot have a weight of –10.5 kilograms.

C If you use bicep girth to predict weight, how accurate do you think the predictions will be?

- Very accurate (typical prediction error will be within a kilogram).
- Somewhat accurate (typical prediction error will be within 10 kilograms).

Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line

- Not very accurate (prediction errors larger than plus or minus 20 kilograms would not be surprising).

> **Answer**: Somewhat accurate.

2 With the following applet, you will investigate how outliers impact the regression line.

http://www.stat.sc.edu/~west/javahtml/Regression.html

A Add points to the scatterplot that are close data points shown. Describe what happens to the regression line.

> **Answer:** The regression line does not change much.

B A data point is *influential* if removing it (or in this case adding it) substantially changes the regression line. Add points to the scatterplot that are outliers relative to the other data points. In other words, add points that are far away from the other data points. Describe what happens to the regression line.

> **Answer:** The regression line changes quite a bit.
>
> **Note:** If you have time and think it is important, the National Council of Teachers of Mathematics website has a nice lesson on investigating the impact of influential points on the slope using an applet: http://illuminations.nctm.org/LessonDetail.aspx?ID=L455.

+++++

This lesson is part of STATWAY™, A Pathway Through College Statistics, which is a product of a Carnegie Networked Improvement Community that seeks to advance student success. Version 1.0, A Pathway Through Statistics, Statway™ was created by the Charles A. Dana Center at the University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching. This version 1.5 and all subsequent versions, result from the continuous improvement efforts of the Carnegie Networked Improvement Community. The network brings together community college faculty and staff, designers, researchers and developers. It is an open-resource research and development community that seeks to harvest the wisdom of its diverse participants in systematic and disciplined inquiries to improve developmental mathematics instruction. For more information on the Statway Networked Improvement Community, please visit **carnegiefoundation.org.** For the most recent version of instructional materials, visit **Statway.org/kernel.**

Lesson 3.2.3
Investigating the Meaning of Numbers in the Equation of a Line

+++++